

Measurement-based Care in Psychiatry—Past, Present, and Future

ABSTRACT

The authors define measurement-based care (MBC) in psychiatry as the use of validated clinical measurement instruments to objectify the assessment, treatment, and clinical outcomes, including efficacy, safety, tolerability, functioning, and quality of life, in patients with psychiatric disorders. MBC includes two processes: routine assessments, such as measuring the severity of symptoms with rating scales, and the use of assessments in decision-making. MBC implementation was tested in the Texas Medication Algorithm Project and the German Algorithm Project and has been shown to improve patient outcomes. Even though more recent research has shown the many benefits of MBC compared to the usual care, MBC is still not the standard of care in psychiatric practice. This review article addresses the advantages of MBC, the barriers to implementing MBC in clinical practice, and the basic properties of MBC instruments. Recent developments in the 21st century that are expected to accelerate the adoption of MBC in clinical practice, including electronic health records, health information technology, and the development of the Standard for Clinicians' Interview in Psychiatry (SCIP) as an MBC tool, will be reviewed. The authors recommend including MBC in psychiatry residency training to promote its use in future generations.

KEYWORDS: Measurement-based care (MBC), Standard for Clinicians' Interview in Psychiatry (SCIP), assessment, psychopathology, assessment tool, rating scale, reliability, validity, outcomes measures, clinical trial

by AHMED ABORAYA, MD, DR.PH; HENRY A. NASRALLAH, MD; DANIEL E. ELSWICK, MD; AHMED ELSHAZLY, MD; NEVINE ESTEPHAN, MD; DALIA ABORAYA, BA; SEHER BERZINGI; JOSLEEN CHUMBERS; SARA BERZINGI, BA; JOHN JUSTICE, MD; JAWAD ZAFAR, DO; and SHEENA DOHAR, MD

Dr. Aboraya is Chief of Psychiatry at William R. Sharpe, Jr. Hospital, Clinical Professor of Psychiatry at West Virginia School of Osteopathic Medicine, and Adjunct Faculty with the School of Public Health West Virginia University, in Weston, West Virginia. Dr. Nasrallah is the Sydney Souers Professor and Chairman of the Department of Psychiatry and Behavioral Neuroscience at Saint Louis University School of Medicine in St. Louis, Missouri. Dr. Elswick is Associate Professor at West Virginia University in Morgantown, West Virginia. Mr. Elshazly is a second year resident at Atlanticare Regional Medical Center in Atlantic City, New Jersey. Dr. Estephan is Child Psychiatry Fellow at West Virginia University in Morgantown, West Virginia. Ms. Aboraya is with Ohio Northern University Law School in Ada, Ohio. Mr. Seher Berzingi and Ms. Chambers and Sara Berzingi are researchers at West Virginia University in Morgantown, West Virginia. Dr. Justice is Medical Director of Sharpe Hospital in Weston, West Virginia. Drs. Zafar and Dohar are psychiatry residents at West Virginia University in Morgantown, West Virginia.

Innov Clin Neurosci. 2018;15(11–12):13–26

In science, measurement is defined as “rules for assigning numbers to objects in such a way as to represent quantities of attributes.”¹ Scientific measurements cannot be valid if they are not reliable. Attributes, reliability, and validity are all crucial to conducting any research. Once scientifically credible measurements are created, testing hypotheses and conducting meaningful clinical trials become possible, leading to advances in science and medicine.

Measurement in psychiatry can be traced back to 1825 when a royal commission was issued to enumerate and measure the “condition of the insane” in the kingdom of Norway. Professor Holst published the results of the survey, which was repeated in 1835 and 1845. The survey results are fascinating and described patients with “mania, melancholia, dementia, idiotia, blind in one eye or two eyes, deaf, dumb, and lepers,” classified by sex and by rural and urban districts.² Major advances in science

are preceded by breakthroughs in measurement methods. This was demonstrated in the field of psychology by the flood of research following the development of intelligence tests and the intelligence quotient (IQ) in 1912.¹

The term *measurement-based care* (MBC) was coined by Trivedi in 2006 and was defined as “the routine measurement of symptoms and side effects at each treatment visit and the use of a treatment manual describing when and how to modify medication doses based on these measures.”³ Other authors had similar definitions: Harding defined MBC as “enhanced precision and consistency in disease assessment, tracking, and treatment to achieve optimal outcomes,”⁴ Arbuckle defined MBC as “a step-by-step approach for assessing, treating, reviewing outcomes and revising treatment in managing medical diseases,”⁵ and Fortney defined MBC as “the systematic administration of symptom rating scales and use of the results to drive clinical decision making at the

FUNDING: No funding was provided.

DISCLOSURES: The authors have no conflicts of interest relevant to the content of this article.

CORRESPONDENCE/REQUEST FOR SCIP MODULES: Ahmed Aboraya, MD, Dr.PH; Email: Aborayascip@gmail.com; Website: www.scip-psychiatry.com

level of the individual patient.”⁶ Our working definition of MBC in psychiatry is “the use of validated clinical measurement instruments to objectify the assessment, treatment and clinical outcomes, including efficacy, safety, tolerability, functioning, and quality of life, in patients with psychiatric disorders.”

MBC refers to two processes: routine assessments, such as measuring the severity of symptoms with rating scales, and the use of assessments in decision-making. The development of rating scales and diagnostic interview schedules during the second half of the 20th century, as well as their use in psychiatric research and clinical trials, was an important catalyst for the development and implementation of MBC. With the publications of *Diagnostic and Statistical Manual of Mental Disorders, Third Edition (DSM-III)* in 1980 and its widespread use worldwide, psychiatric research and clinical trials flourished as geneticists, pharmacologists, and neuroscientists became research partners with investigative psychiatrists.⁷ More clinical trials were conducted to assess the efficacy and safety of the new psychotropic medications all over the world.^{8–16} With the availability of rating scales and standardized diagnostic interviews, the Texas Medication Algorithm Project (TMAP) and the German Algorithm Project (GAP) tested the implementation of MBC in outpatient and inpatient clinical settings and have shown that MBC can positively impact patient outcomes.^{17–19} Even though the term *measurement-based care* is relatively new in psychiatric literature, it has been an integral component of randomized, clinical trials for decades.²⁰

The other popular and common method of caring for patients is the “standard” or “usual” care that has been provided by clinicians daily for centuries. *Usual standard care* (USC) for patients involves the same two components of MBC: assessment and decision-making. Clinicians, by training, assess psychopathology and its severity and make decisions based on their assessment, without using rating scales or standardized diagnostic interviews. In 1933, Hardcastle et al studied the present condition of the first 100 patients (adults and children) who attended the Department of Psychological Medicine at Guy’s Hospital in London in 1931. Although clinicians in 1933 did not have or use the Hamilton Depression Rating Scale (HAM-D)

or other scales we have today, they evaluated the patients and grouped them into four main groups: much improved, improved, unchanged, or worse. Based on their evaluations, they made decisions to admit or treat patients accordingly. The Hardcastle study was published in the *Journal of Mental Science* in 1934.²¹ In the same journal and during the same year, Lewis²² published a 102-page monograph describing in great detail the symptoms and signs of 61 cases of “depressive state,” all examined and treated by Lewis between the years 1928 and 1929 in the Maudsley Hospital in London, England. One might make the case that psychiatrists at Guy’s and Maudsley’s hospitals in 1934 had more expertise in psychopathology assessment than today’s psychiatrists because one of the unintended consequences of the *DSM* era is the limitation of psychopathology training according to *DSM* and *International Statistical Classification of Diseases and Related Health Problems (ICD)* criteria.²³

Recent research has shown the superiority of MBC compared to USC in improving patient outcomes.^{6,24–26} A recent, well-designed, blind-rater, randomized trial by Guo et al¹⁷ showed that MBC, *per se*, is more effective than USC in achieving response and remission and lowering the time to response and remission. Given the evidence of the benefits of MBC in improving patient outcomes, an important question arises: Why has MBC not yet been established as the standard of care in clinical practice?

This review article addresses the advantages of MBC, the barriers to implementing MBC in clinical practice, and important contemporary developments in the 21st century that are expected to accelerate the adoption of MBC in clinical practice.

ADVANTAGES OF MBC

Research over the past 20 years has shown that MBC improves the quality of patient care, and leaders in the mental health field have been calling for the integration of MBC into routine care.⁶ Compared to the usual care, MBC has been shown to do the following:

1. Improve psychotherapy outcomes⁶
2. Monitor symptom reduction in patients with psychiatric disorders, such as anxiety, depression, and bipolar^{27–29}
3. Identify patients who are improving and those who are deteriorating^{6,30,31}
4. Improve role functioning, satisfaction

with care, quality of care, and quality of life^{24,29,32}

5. Enhance the therapeutic relationship and communication between providers and patients⁶
6. Improve collaborative care efforts among providers^{24,32}
7. Improve the accuracy of clinical judgment^{4,33}
8. Close the gap between research and practice, and move psychiatry into the mainstream of medicine⁴
9. Enhance the clinician’s decision-making process^{24,26}
10. Enhance individualized treatment³⁴
11. Be transdiagnostic and transtheoretical²⁴
12. Be feasible to implement on a large scale^{3,35–38}

BARRIERS TO MBC

Even though recent research has shown the many benefits of MBC compared to USC, MBC is still not the standard of care in clinical settings, and a small proportion of clinicians use outcome assessments.^{4,39} Many psychiatric measures with good psychometric properties have been developed and tested over the past decades (e.g., standardized diagnostic interviews, rating scales, and self-rating scales).^{40–54} However, most of these measures are used in research and clinical trials and not in clinical settings. A study by Hatfield²⁶ reported that 37.1 percent of clinicians use some form of outcome assessments, and 62.9 percent do not use any outcome measures. Zimmerman⁵⁵ reported that more than 80 percent of psychiatrists indicated they did not routinely use scales to monitor outcome when treating depression. In a survey of psychiatric practitioners, Nasrallah⁵⁶ reported that 98 percent of psychiatrists indicated they do not use any of the four clinical rating scales routinely used in clinical trials and are required for the United States Food and Drug Administration (FDA) approval of psychiatric medications. These four scales are 1) Positive and Negative Syndrome Scale (PANSS), 2) Young Mania Rating Scale (YMRS), 3) HAM-D, and 4) Montgomery-Asberg Depression Rating Scale (MADRS). The vast majority of the surveyed participants attributed their avoidance of rating scales to “lack of time.” Many other authors have noted that clinicians

do not use standardized scales in clinical practice.^{57–63} Barriers to implementing MBC are summarized in Table 1.

Additionally, theoretical orientation was described as a potential barrier for insight-oriented therapists, who were less likely than cognitive or behavioral therapists to use outcome measures.³⁹ However, a recent article by Scott²⁴ demonstrated that clinicians can implement MBC regardless of their theoretical orientation or training background.

IMPLEMENTATION OF MBC

To encourage clinicians to use measures in clinical care decisions, measures should have the following basic properties:

1. Efficient (Measures should be brief and not time-consuming to the clinician.^{4,67} A rating scale completed by the clinician should take no more than a few minutes to administer.)
2. Established as reliable and valid⁴
3. User-friendly and a reflection of what clinicians do in clinical settings⁶⁷
4. Brief (Self-rating scales completed by patients should take no more 2–3 minutes to complete) and simple (Directions should be easy to follow to improve patient willingness to take the test at each follow up visit.)⁶⁸
5. Clinically meaningful and useful, covering the criteria and symptom domains of the disorder⁶⁷
6. Clinically relevant to decision-making⁶⁵
7. Easily extractable and not embedded in progress notes⁶
8. Sensitive to changes induced by medications or psychotherapy.⁶⁹

DEVELOPMENT OF THE STANDARD FOR CLINICIANS' INTERVIEW IN PSYCHIATRY (SCIP) AND THE SCIP SCALES AS AN MBC TOOL

After Ahmed Aboraya (first author) finished his master's and doctoral degrees at Johns Hopkins University in 1991, he started his psychiatry residency training with a determination to use psychiatric measures in clinical settings. Disappointed after 10 years of trying to use almost all of the relevant existing scales and standardized diagnostic interviews for adult psychiatric disorders, Aboraya concluded that existing measures were not practical for use in the real world of psychiatric

TABLE 1. Barriers to measurement-based care (MBC)

1. Measures are time consuming (most commonly cited reason by psychiatrists) ^{55,56,61}
2. Measures are designed for research use and not for clinician use ^{56,63}
3. Ratings produced by measures might not always be clinically relevant ^{64,65}
4. Administering rating scales might interfere with establishing rapport with patients ⁶⁶
5. The perception that measures are not more useful than clinical assessment ^{55,66}
6. The perception that MBC is over-systematizing and depersonalizing ⁴
7. Some measures, such as standardized diagnostic interviews, can be cumbersome, unwieldy, and complicated ⁶⁴
8. Cost and lack of resources to implement MBC ²⁶
9. Limited formal training (included in top two barriers for residents and faculty) ^{26,66}
10. Lack of protocols and training manuals ²⁴
11. Lack of consensus as to which instrument to use for a given disorder ⁶⁶
12. Absence of a requirement to use MBC—few work settings require MBC ^{26,66}
13. Lack of incentives to use MBC
14. Complexity of patients with multiple overlapping comorbidities
15. The perception that measures “restrict the flexibility and creativity” of the interviewer

practice. Consequently, he embarked on developing the Standard for Clinicians' Interview in Psychiatry (SCIP) as a tool for clinicians in real clinical settings for assessment and decisions-making. In other words, the SCIP was designed from the outset as an MBC tool. The SCIP was tested in an international, multisite study in three countries (United States, Canada, and Egypt) between the years 2000 and 2012. The total sample size, including all sites, was 1,004 subjects, making the SCIP project the largest validity and reliability study to be conducted on diagnostic interviews in psychiatry.^{47,48,64} The details of the design of the SCIP project were published in 2014.⁴⁸ In addition to being the only tool designed from the outset for use in MBC, the SCIP has two unique advantages: the development of comprehensive and reliable items measuring psychopathology and the creation of reliable and validated SCIP scales for adult psychiatric disorders.

The development of reliable psychopathology items. Inter-rater reliability (Kappa) of the SCIP was measured on 150 items covering anxiety, panic, obsessive compulsive disorder (OCD), posttraumatic stress disorder (PTSD), depression, mania, psychosis, disorganized behavior, negative symptoms, alcohol, and drug psychopathology domains.⁴⁸ To calculate stable Kappa for attention deficit hyperactivity disorder (ADHD) and eating disorders, an additional 40 young and predominantly female patients were interviewed at William R. Sharpe, Jr. Hospital and Chestnut Ridge Center by at least two

interviewers at the same time (to establish inter-rater reliability). The mean patient age was 35, with 68 percent being female, 90 percent being white, and 73 percent completing at least 12 years of education. If the patient was interviewed by three interviewers (i.e., A, B, C), a comparison was made between interviewer A and B, A and C, and B and C. A total of 75 comparisons allowed the calculation of stable Kappa for ADHD and eating disorders. Table 2 shows inter-rater reliability agreement (Kappa) and the standard error for 206 psychopathology items based on the interviews of 322 patients from William R. Sharpe Jr. Hospital, Chestnut Ridge Center (inpatient and outpatient), Ain Shams University Hospital, and Mansoura University Hospital. The mean patient age was 33, with 45 percent being female, 97 percent being white, and 63 percent completing at least 12 years of education. Five items (Item Numbers 102, 104, 167, 184, and 186) had unstable Kappa, and 201 items had stable Kappa. Out of 201 items with stable Kappa, 165 items (82%) had satisfactory agreement ($\kappa > 0.7$), 30 items (15%) had fair agreement ($\kappa = 0.5$ to 0.7), and 6 items (3%) had poor agreement ($\kappa < 0.5$).

In 1992, Nancy Andreasen, a renowned researcher, stressed the importance of establishing reliability at the level of individual symptoms and signs. Creating reliable psychological dimensions requires reliability of the items measuring individual symptoms and signs. The absence of valid and reliable symptoms was the main limiting factor in creating dimensional measures in the past.⁷⁰

TABLE 2. Inter-rater reliability agreement (kappa) and standard error (SE) for the Standard for Clinicians' Interview in Psychiatry (SCIP) items (symptoms and signs) in patients at William R. Sharpe Jr. Hospital, Chestnut Ridge Center (inpatient and outpatient), Ain Shams University Hospital, and Mansoura University Hospital.

ITEM #		SCIP ITEMS	TOTAL # OF POSITIVE CASES FOR A GIVEN ITEM	KAPPA (*)	SE
1	1	Generalized anxiety	61	0.76	0.05
2	2	Panic attacks	54	0.81	0.05
3	3	Agoraphobia	26	0.52	0.05
4	4	Social phobia	22	0.51	0.05
5	5	Screening for obsessions	38	0.70	0.04
6	6	Screening for compulsions	31	0.58	0.05
7	7	Witness or experience traumatic events	69	0.75	0.05
8	8	Re-experience traumatic events	34	0.89	0.05
9	9	Depressed mood	158	0.86	0.04
10	10	Anhedonia	125	0.87	0.04
11	11	Suicidal ideation, intention, plan	79	0.61	0.04
12	12	Elated mood	76	0.72	0.05
13	13	Irritable mood	65	0.75	0.05
14	14	Mixed mood (same day mood changes)	44	0.50	0.05
15	15	Paranoid delusions	97	0.83	0.04
16	16	Other delusions	39	0.77	0.04
17	17	Auditory hallucinations	92	0.76	0.04
18	18	Other hallucinations	51	0.68	0.05
19	19	Violence	74	0.64	0.04
20	20	Disorganized behavior	32	0.54	0.04
21	21	Disorganized thoughts	39	0.65	0.04
22	22	Alcohol problems	53	0.89	0.06
23	23	Drug problems	17	0.78	0.06
24	24	Somatic symptoms	33	0.81	0.05
25	25	Pain symptoms	24	0.93	0.05
26	26	Worry about weight and image	12	0.73	0.05
27	27	Binge eating	27	0.97	0.12
28	28	Poor attention	11	0.73	0.05
29	29	Hyperactivity	14	0.58	0.05
30	1	Panic attacks	30	0.92	0.06
31	2	Worry about having another panic attack	25	0.81	0.04
32	3	Action to prevent panic attacks	26	0.87	0.04
33	4	Generalized anxiety	25	0.84	0.04
34	5	Restlessness with anxiety	26	0.74	0.04
35	6	Tension with anxiety	22	0.77	0.04
36	7	Exhaustion with anxiety	22	0.79	0.05
37	8	Poor concentration with anxiety	27	0.76	0.05
38	9	Irritability with anxiety	28	0.83	0.04
39	10	Insomnia with anxiety	25	0.82	0.05
40	11	Obsessions	26	0.85	0.04
41	12	Compulsions	18	0.77	0.04
42	1	Experienced traumatic events	10	0.83	0.05
43	2	Distressing recollection of events	30	0.88	0.05
44	3	Bad dreams or nightmares	26	0.94	0.05
45	4	Flashback	23	0.87	0.05
46	5	Psychological distress due to events	26	0.91	0.05

The SCIP study removed this major obstacle by creating reliable symptoms and signs for 206 psychopathology items, which paved the way for the creation of reliable and valid SCIP dimensions and scales.

The development of reliable and valid SCIP dimensions and scales for adult psychiatric disorders. The SCIP dimensions and scales were created based on the interviews of 700 patients, 670 of whom were from William R. Sharpe Jr. Hospital in Weston, West Virginia, and 30 of whom were from at Chestnut Ridge Center in Morgantown, West Virginia. Mean patient age was 34, with 59 percent being male, 95 percent being white, and 66 percent completing at least 12 years of education. Patients were evaluated and diagnosed by the attending psychiatrist. We evaluated and treated each patient from admission to discharge, using all available data, including information from previous hospitalizations and family members, labs, psychological testing, and diagnostic schedules, as needed, to reach the final diagnoses.

The initial items of the SCIP dimensions were formulated based upon the *DSM* and *ICD* criteria and expert opinions. The sensitivity and specificity of the initial dimensions were calculated against the psychiatric diagnosis, as described above. Rules for shortening the lengthy initial dimensions and creating the final SCIP dimensions included removing items with low prevalence, low sensitivity, or low item-rest correlation (<0.4). The reliability and validity of the remaining items were recalculated with repetitive iterations. The sensitivity and specificity of the final dimensions were approximately equal to the sensitivity and specificity of the initial dimensions. Appendix I shows the initial depression dimension, which has 15 symptoms and signs of depression. Three items not covered in *DSM-5*—crying when depressed, feeling hopeless, and reduced sexual desire—were included in the initial depression dimension based on the recommendations and use by experts and clinicians for decades, even before the existence of the *DSM*.²³ The sensitivity and specificity of the initial depression scale were 93.24 percent and 74.15 percent, respectively. Following the rules of creating the SCIP scales, the final core depression scale had eight items with 93.24-percent sensitivity and 72.32-percent specificity.

Based upon reliable psychopathology items, the SCIP is the only diagnostic tool that has 18 inherent rating scales for the following

domains: generalized anxiety, obsessions, compulsion, posttraumatic stress, depression, mania, delusions, hallucinations, disorganized thoughts, aggression, negative symptoms, alcohol use, drug use, attention deficit, hyperactivity, anorexia, binge-eating, and bulimia. Each of the SCIP rating scales takes 2 to 5 minutes to complete. The SCIP rating scales meet the criteria for MBC because they are efficient, reliable, valid, reflect how clinicians assess psychiatric disorders, and are relevant to decision-making. These unique properties make the SCIP ideal as an MBC tool. Tables 3 to 15 show the items included in the SCIP scales, item rest correlation, mean interitem correlation, Cronbach's alpha with one-sided 95-percent confidence interval (CI), sensitivity and specificity at the optimal cutpoint, and receiver operating characteristic (ROC) area with standard error. All of the SCIP scales have been validated with the exception of the OCD and eating disorders scales. Aboraya, Henry Nasrallah, and Daniel Elswick (the first three authors of this article) are currently writing a book that describes the advantages and disadvantages of the SCIP scales and other existing scales in the literature.

RECENT DEVELOPMENTS AFFECTING MBC

Electronic health records. Electronic health records (EHR) are being used across clinical settings, from big academic institutions to solo practices. The United States Federal government has given financial incentives to solo practitioners to use EHR, and most academic institutions use advanced EHR.⁷¹ Once MBC tools are identified, they can be uploaded to the EHR and be readily available for clinicians to use. The use of EHR should facilitate the implementation of MBC.⁴

Health information technology. Advances in health information technology, such as software programs, handheld devices, web-based training, and videos, should facilitate clinician training and use of MBC tools.^{6,71,72} Currently, psychiatrists record diagnosis, mental status, and other clinical aspects in a loose narrative outline, making it difficult to measure or compare outcomes of patients that have been assessed by different clinicians.⁶⁷ This current practice will be outdated in the near future with the implementation of MBC. With the right software and integrated EHR, clinicians

TABLE 2, continued. Inter-rater reliability agreement (kappa) and standard error (SE) for the Standard for Clinicians' Interview in Psychiatry (SCIP) items (symptoms and signs) in patients at William R. Sharpe Jr. Hospital, Chestnut Ridge Center (inpatient and outpatient), Ain Shams University Hospital, and Mansoura University Hospital.

ITEM #		SCIP ITEMS	TOTAL # OF POSITIVE CASES FOR A GIVEN ITEM	KAPPA (*)	SE
47	6	Physical reactions due to events	24	0.93	0.05
48	7	Avoidance of thoughts and feelings	27	0.94	0.05
49	8	Avoidance of people, places	27	0.94	0.05
50	9	Amnesia	15	0.70	0.06
51	10	Diminished interest	17	0.83	0.05
52	11	Detachment and isolation	22	0.87	0.05
53	12	Diminished affect	24	0.88	0.05
54	13	Insomnia	16	0.78	0.05
55	14	Anger	19	0.80	0.05
56	15	Poor concentration	14	0.78	0.05
57	16	Hypervigilance	17	0.87	0.05
58	17	Startle response	20	0.86	0.05
59	18	Daze (feeling out of touch with surroundings)	16	0.82	0.05
60	1	Depressed mood	128	0.91	0.04
61	2	Anhedonia	121	0.87	0.04
62	3	Crying when depressed	11	0.76	0.04
63	4	Hopelessness	11	0.82	0.04
64	5	Fatigue and loss of energy	97	0.72	0.04
65	6	Poor concentration	116	0.80	0.04
66	7	Psychomotor retardation	97	0.72	0.04
67	8	Appetite changes when depressed	93	0.79	0.04
68	9	Weight loss	62	0.71	0.04
69	10	Weight gain	15	0.76	0.05
70	11	Initial insomnia	103	0.79	0.04
71	12	Middle insomnia	79	0.65	0.04
72	13	Late insomnia	46	0.62	0.04
73	14	Hypersomnia	26	0.68	0.05
74	15	Decreased libido	74	0.80	0.04
75	16	Worthlessness	97	0.78	0.04
76	17	Guilt	86	0.80	0.04
77	18	Suicide	68	0.64	0.04
78	1	Elated mood	71	0.75	0.04
79	2	Irritable mood	70	0.76	0.04
80	3	Mixed mood (same day mood changes)	41	0.58	0.05
81	4	Racing thoughts	71	0.85	0.04
82	5	Pressured speech	53	0.72	0.04
83	6	Flight of ideas	15	0.62	0.06
84	7	Clanging	12	0.49	0.04
85	8	Distraction	63	0.79	0.04
86	9	Increase in activities	68	0.83	0.04
87	10	Grandiosity	40	0.81	0.04
88	11	Impulsivity	41	0.92	0.12
89	12	Over spending	49	0.74	0.04
90	13	Decreased sleep	56	0.78	0.04
91	14	Hypersexuality	24	0.69	0.04
92	1	Auditory hallucinations	54	0.90	0.04

TABLE 2, continued. Inter-rater reliability agreement (kappa) and standard error (SE) for the Standard for Clinicians' Interview in Psychiatry (SCIP) items (symptoms and signs) in patients at William R. Sharpe Jr. Hospital, Chestnut Ridge Center (inpatient and outpatient), Ain Shams University Hospital, and Mansoura University Hospital.

ITEM #		SCIP ITEMS	TOTAL # OF POSITIVE CASES FOR A GIVEN ITEM	KAPPA (*)	SE
93	2	Hallucinations frequency	54	0.93	0.05
94	3	Internal hallucinations	50	0.84	0.04
95	4	Voices commenting	40	0.77	0.04
96	5	Second and third hallucinations	45	0.78	0.04
97	6	Visual hallucinations	27	0.81	0.04
98	7	Other hallucinations	10	0.95	0.05
99	8	Observed hallucinations	12	0.55	0.04
100	9	Reading thoughts	17	0.83	0.04
101	10	Thought insertion	16	0.76	0.04
102	11	Thought withdrawal	6	0.8 (**)	0.04
103	12	Thought broadcasting	16	0.71	0.04
104	13	Somatic passivity	7	0.58 (**)	0.04
105	14	Paranoid delusions	50	0.86	0.04
106	15	Conspiracy delusions	49	0.84	0.04
107	16	Delusions of reference	31	0.81	0.05
108	17	Religious delusions	17	0.80	0.04
109	18	Grandiose delusions	16	0.77	0.05
110	19	Other delusions	12	0.40	0.05
111	20	Bizarreness of delusions	14	0.43	0.05
112	1	Derailment	37	0.65	0.06
113	2	Flight of ideas	15	0.62	0.06
114	3	Tangentiality	28	0.57	0.06
115	4	Incoherent speech	18	0.41	0.06
116	5	Illogical speech	13	0.25	0.05
117	1	Agitation	33	0.48	0.04
118	2	Violence	25	0.64	0.04
119	3	Odd behavior	19	0.67	0.06
120	4	Inappropriate affect	14	0.77	0.06
121	1	Alogia	29	0.62	0.05
122	2	Anhedonia	121	0.87	0.04
123	3	Affective flattening or blunting	42	0.68	0.05
124	4	Avolition	35	0.74	0.04
125	5	Asociality	35	0.74	0.04
126	6	Attention impairment	41	0.92	0.12
127	7	Psychomotor slowing	97	0.72	0.04
128	8	Poor self care	27	0.79	0.06
129	1	Alcohol tolerance	39	0.99	0.06
130	2	Alcohol withdrawal	33	0.93	0.06
131	3	Drinking alcohol to avoid withdrawal	29	0.96	0.06
132	4	Unable to control alcohol	51	0.96	0.06
133	5	Unable to reduce or stop alcohol	47	0.85	0.06
134	6	Time spent to drink alcohol	37	0.94	0.06
135	7	Failure to fulfil major obligations	36	0.92	0.06
136	8	Giving up social or recreational activities	36	0.92	0.06
137	9	Fighting when intoxicated	31	0.90	0.06
138	10	Alcohol family problems	51	0.82	0.06

should be able to efficiently record a rating scale, calculate the scale score, compare scores on the same scale over time, draw graphs, and do analyses.

Training psychiatry residents and clinicians in MBC. Lack of training was listed among the top two barriers to using MBC by psychiatry residents and faculty.^{26,66} In addition, lack of consensus as to which instrument to use was another barrier due to the availability of many measures.⁶⁶ One important place to promote the use of MBC is in psychiatry residency programs. Currently, no specific requirements exist to evaluate training on the use of MBC during residency.⁷³ A new psychiatry subcompetency for MBC could be added to the existing 22 psychiatry subcompetencies included in the Psychiatry Milestone Project Initiative by the Accreditation Council for Graduate Medical Education (ACGME) and American Board of Psychiatry and Neurology (ABPN).⁷⁴ Psychiatry residents could learn and progress using the new MBC subcompetency from Level 1 (basic knowledge of psychiatric measures) to Levels 4 and 5 (the ability to use the appropriate measures for making decisions). Arbuckle et al⁵ implemented a curriculum in MBC for depression in a psychiatric resident clinic and found that MBC was feasible and improved depression screening and monitoring. Aboraya is developing an MBC manual and a didactic seminar for psychiatry residents, using the SCIP scales and other scales for personality disorders and cognitive disorders. A pilot study for implementing MBC for adult psychiatric disorders at the West Virginia University residency program and other programs is underway. If psychiatry residents are trained in MBC, they might potentially practice MBC for the rest of their careers. There is also urgent need to train faculty and clinicians in MBC through continuing medical education (CME) workshops.⁴ Aboraya, Nasrallah, and Elswick are planning MBC workshops to train clinicians and psychiatry residents on how to choose the right scale or instrument for each individual patient.

DISCUSSION

In 1961, when Robert Spitzer developed the Mental Status Schedule, the first published structured interview in the United States,⁷⁵ the *New York Post* published an article in 1963 that stated "a young doctor at Columbia University's New York State Psychiatric Institute

has developed a tool which may become the psychiatrist's thermometer and microscope and X-ray machine rolled into one."⁷⁶ Five decades later, many might say this statement is still accurate—measures in psychiatry could be considered the equivalent of a thermometer and a stethoscope to a physician. No measure, scale, or diagnostic interview will ever replace a seasoned, experienced clinician who has been evaluating and treating real patients for years. MBC is not intended to replace clinical judgment and cannot substitute for an observant and caring clinician.⁴ Just as thermometers, stethoscopes, and lab tests help other types of physicians reach accurate diagnoses and provide appropriate management, the use of MBC by psychiatrists has the potential to improve the accuracy of diagnoses and improve the outcomes of care. In essence, MBC aims to get the diagnosis and management right as often and as quickly as possible.⁴

The use of scientific rules and expert input for the creation of efficient and validated SCIP scales does not minimize the importance of the psychopathology items not included in the final SCIP scales. The core depression scale of the SCIP does not include questions on reduced sexual drive, sleep, or appetite changes. Clinicians need to inquire about these important items because they can impact which medications will be most effective for individual patients. In teaching and implementing MBC, clinicians should stress the importance of comprehensive psychopathological assessment to avoid the trap of limiting psychopathology education to specific diagnostic criteria or certain scales.

CONCLUSION

Recent studies have shown that the cost of MBC implementation is minimal and the benefits are significant for patients, providers, and payers.⁶ The advantages of MBC outweigh the challenges to its implementation.⁷⁷ Moreover, many payers and accreditation organizations are requiring the use of MBC in psychiatric practice. We believe it is better for healthcare providers to develop their own MBC tools than to have outcome measures imposed on them by payers and/or regulators.⁶ The three main ingredients for MBC implementation, namely measures, EHR, and health information technologies, already exist. We believe now is the time to employ MBC into standard practice, and published research supports this.²⁰ The onus lies on mental health providers to implement MBC.

TABLE 2, continued. Inter-rater reliability agreement (kappa) and standard error (SE) for the Standard for Clinicians' Interview in Psychiatry (SCIP) items (symptoms and signs) in patients at William R. Sharpe Jr. Hospital, Chestnut Ridge Center (inpatient and outpatient), Ain Shams University Hospital, and Mansoura University Hospital.

ITEM #		SCIP ITEMS	TOTAL # OF POSITIVE CASES FOR A GIVEN ITEM	KAPPA (*)	SE
139	11	Alcohol legal problems	29	0.92	0.06
140	12	Alcohol medical problems	11	0.70	0.06
141	13	Continue alcohol with problems	57	0.87	0.06
142	14	Alcohol in hazardous situations	42	0.77	0.06
143	15	Alcohol binge	37	0.88	0.06
144	16	Alcohol blackout	53	0.98	0.06
145	1	Drug tolerance	49	0.95	0.06
146	2	Drug withdrawal	46	0.97	0.06
147	3	Using drug to avoid withdrawal	40	0.94	0.06
148	4	Unable to control drug use	55	0.97	0.06
149	5	Unable to reduce or stop drug use	54	0.97	0.06
150	6	Time spent to use drug	56	0.88	0.06
151	7	Failure to fulfil major obligations	50	0.95	0.06
152	8	Giving up social or recreational activities	50	0.95	0.06
153	9	Fighting when using drug	22	0.80	0.06
154	10	Drug family problems	58	0.80	0.06
155	11	Drug legal problems	22	0.80	0.06
156	12	Drug emotional problems	19	0.76	0.06
157	13	Drug use with problems	64	0.91	0.06
158	14	Drug use in hazardous situations	57	0.90	0.06
159	1	Being underweight	32	0.83	0.11
160	2	Weight affect feelings	50	0.75	0.12
161	3	Fear of weight gain	20	1.00	0.12
162	4	Losing weight by fasting	32	0.95	0.12
163	5	Losing weight by exercise	22	0.86	0.12
164	6	Losing weight by diet pills	22	0.97	0.12
165	7	Losing weight by vomiting	27	0.94	0.12
166	8	Losing weight by laxatives	14	1.00	0.12
167	9	Losing weight by other methods	8	1.00 (**)	0.12
168	10	Binge eating	27	0.97	0.12
169	11	Binge eating frequency	27	0.85	0.09
170	12	Losing control with binge eating	17	0.96	0.12
171	13	Binge eating behavior	27	1.00	0.12
172	14	Eating fast during binge eating	16	1.00	0.12
173	15	Eating until uncomfortably full during binge eating	25	0.94	0.12
174	16	Eating when not hungry	22	0.97	0.12
175	17	Eating alone	16	0.96	0.12
176	18	Feeling disgusted and guilty	22	0.86	0.12
177	19	Distressed by overeating	24	0.77	0.11
178	20	Compensatory behavior after binge eating	25	0.97	0.12
179	21	Fasting after binge eating	19	0.93	0.12
180	22	Exercise after binge eating	12	0.95	0.12
181	23	Using diet pills after binge eating	12	0.95	0.12
182	24	Vomiting after binge eating	17	1.00	0.12
183	25	Taking laxatives after binge eating	14	1.00	0.12

TABLE 2, continued. Inter-rater reliability agreement (kappa) and standard error (SE) for the Standard for Clinicians' Interview in Psychiatry (SCIP) items (symptoms and signs) in patients at William R. Sharpe Jr. Hospital, Chestnut Ridge Center (inpatient and outpatient), Ain Shams University Hospital, and Mansoura University Hospital.

ITEM #	SCIP ITEMS	TOTAL # OF POSITIVE CASES FOR A GIVEN ITEM	KAPPA (*)	SE
184	26 Other losing weight methods after binge eating	9	1.00 (**)	0.12
185	27 Binge eating compensatory behavior frequency	25	0.87	0.09
186	28 Other eating behaviors	4	0.39 (**)	0.09
187	1 Attention difficulty	41	0.92	0.12
188	2 Long attention difficulty	39	0.95	0.12
189	3 Avoiding tasks	34	0.97	0.12
190	4 Attention when spoken to	32	0.97	0.12
191	5 Organization and meeting deadlines	30	0.82	0.12
192	6 Changing activities	40	0.92	0.12
193	7 Distraction	43	0.97	0.12
194	8 Misplacing things	43	0.94	0.12
195	9 Forgetting daily activities	24	0.94	0.12
196	10 Losing track	40	0.92	0.12
197	11 Fidgety	41	0.81	0.12
198	12 Leaving seats	30	0.88	0.12
199	13 Restlessness/moving	49	0.61	0.12
200	14 Hyperactivity	22	0.97	0.12
201	15 Waiting in line	23	1.00	0.12
202	16 Talking too much	12	1.00	0.12
203	17 Loud and noisy	22	0.58	0.11
204	18 Impulsivity	41	0.92	0.12
205	19 Disturbing others	23	0.97	0.12
206	20 Blurt out answers	32	0.89	0.12

* Kappa values were calculated based upon inter-rater interviews of 322 patients at William R. Sharpe Jr. Hospital, Chestnut Ridge Center (inpatient and outpatient), Ain Shams University Hospital and Mansoura University Hospital.

**Kappa is unstable if the number of positive cases for a given item is <10.

REFERENCES

- Nunnally JC. *Psychometric Theory*. New York McGraw-Hill. 1978.
- Holst P. On the statistics of the insane, blind, deaf and dumb, and lepers, of Norway. *J Royal Stat Soc*. 1852;15:250–256.
- Trivedi MH, Rush AJ, Wisniewski SR, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006;163:28–40.
- Harding KJ, Rush AJ, Arbuckle M, et al. Measurement-based care in psychiatric practice: a policy framework for implementation. *J Clin Psychiatry*. 2011;72:1136–1143.
- Arbuckle MR, Weinberg M, Kistler SC, et al. A curriculum in measurement-based care: screening and monitoring of depression in a psychiatric resident clinic. *Acad Psychiatry*. 2013;37:317–320.
- Fortney JC, Unutzer J, Wrenn G, et al. A tipping point for measurement-based care. *Psychiatr Serv*. 2016;appips201500439.
- McHugh PR. Striving for coherence: psychiatry's efforts over classification. *JAMA*. 2005;293:2526–2528.
- Fleischhacker WW, Gopal S, Lane R, et al. A randomized trial of paliperidone palmitate and risperidone long-acting injectable in schizophrenia. *Int J Neuropsychopharmacol*. 2012;15:107–118.
- Fleischhacker WW, McQuade RD, Marcus RN, et al. A double-blind, randomized comparative study of aripiprazole and olanzapine in patients with schizophrenia. *Biol Psychiatry*. 2009;65:510–517.
- Kane JM, Eerdekens M, Lindenmayer JP, et al. Long-acting injectable risperidone: efficacy and safety of the first long-acting atypical antipsychotic. *Am J Psychiatry*. 2003;160:1125–1132.
- Kane JM, Skuban A, Ouyang J, et al. A multicenter, randomized, double-blind, controlled phase 3 trial of fixed-dose brexpiprazole for the treatment of adults with acute schizophrenia. *Schizophr Res*. 2015;164:127–135.
- Kane JM, Zukin S, Wang Y, et al. Efficacy and safety of cariprazine in acute exacerbation of schizophrenia: results from an international, Phase III clinical trial. *J Clin Psychopharmacol*. 2015;35:367–373.
- Nasrallah HA, Silva R, Phillips D, et al. Lurasidone for the treatment of acutely psychotic patients with schizophrenia: a 6-week, randomized, placebo-controlled study. *J Psychiatr Res*. 2013;47:670–677.
- Durgam S, Cutler AJ, Lu K, et al. Cariprazine in acute exacerbation of schizophrenia: a fixed-dose, phase 3, randomized, double-blind, placebo- and active-controlled trial. *J Clin Psychiatry*. 2015;76:e1574–1582.
- Durgam S, Starace A, Li D, et al. An evaluation of the safety and efficacy of cariprazine in patients with acute exacerbation of schizophrenia: a phase II, randomized clinical trial. *Schizophr Res*. 2014;152:450–457.
- Meltzer HY, Risinger R, Nasrallah HA, et al. A randomized, double-blind, placebo-controlled trial of aripiprazole lauroxil in acute exacerbation of schizophrenia. *J Clin Psychiatry*. 2015;76:1085–1090.
- Guo T, Xiang YT, Xiao L, et al. Measurement-based care versus standard care for major depression: a randomized controlled trial with blind raters. *Am J Psychiatry*. 2015;172:1004–1013.
- Crismon ML, Trivedi M, Pigott TA, et al. The Texas Medication Algorithm Project: report of the Texas Consensus Conference Panel on Medication Treatment of Major Depressive Disorder. *J Clin Psychiatry*. 1999;60:142–156.
- Ricken R, Wiethoff K, Reinhold T, et al. Algorithm-guided treatment of depression reduces treatment costs—results from the randomized controlled German Algorithm Project (GAPIL). *J Affect Disord*. 2011;134:249–256.
- Rush AJ. Isn't it about time to employ measurement-based care in practice? *Am J Psychiatry*. 2015;172:934–936.

TABLE 3. The items of the Standard for Clinicians' Interview in Psychiatry (SCIP) generalized anxiety scale, item rest correlation, mean interitem correlation, Cronbach's alpha with one-sided 95% confidence interval (CI), sensitivity and specificity at the optimal cutpoint, and receiver operating characteristic (ROC) area with standard error (SE). Data based upon 700 patients interviewed at Sharpe Hospital and Chestnut Ridge Center.

GENERALIZED ANXIETY SCALE ITEMS	ITEM REST CORRELATION	MEAN INTERITEM CORRELATION	CRONBACH'S ALPHA (ONE-SIDED 95% CI)	VALIDITY AT CUTPOINT	SENSITIVITY	SPECIFICITY	ROC AREA (SE)
1. Anxiety	0.2854	0.6774	0.9363 (≥ 0.9301)	≥ 2	77.78%	97.76%	0.9889 (0.0036)
2. Restlessness with anxiety	0.8957						
3. Tension with anxiety	0.9121						
4. Exhaustion with anxiety	0.8670						
5. Poor concentration with anxiety	0.8926						
6. Irritability with anxiety	0.8485						
7. Insomnia with anxiety	0.9027						

TABLE 4. The items of the Standard for Clinicians' Interview in Psychiatry (SCIP) core posttraumatic stress disorder (PTSD) scale, item rest correlation, mean interitem correlation, Cronbach's alpha with one-sided 95% confidence interval (CI), sensitivity and specificity at the optimal cutpoint, and receiver operating characteristic (ROC) area with standard error (SE). Data based upon 700 patients interviewed at Sharpe Hospital and Chestnut Ridge Center.

CORE PTSD SCALE ITEMS	ITEM REST CORRELATION	MEAN INTERITEM CORRELATION	CRONBACH'S ALPHA (ONE-SIDED 95% CI)	VALIDITY AT CUTPOINT	SENSITIVITY	SPECIFICITY	ROC AREA (SE)
1. Traumatic experience	0.6695	0.6403	0.9586 (≥ 0.9547)	≥ 4	93.75%	98.42%	0.9868 (0.0082)
2. Distressing memories	0.8618						
3. Nightmares/bad dreams	0.8354						
4. Flashback	0.8222						
5. Avoidance	0.8599						
6. Amnesia	0.6080						
7. Diminished interest	0.7384						
8. Detached/distant	0.8118						
9. Diminished affect	0.8313						
10. Insomnia	0.8001						
11. Anger	0.7598						
12. Hypervigilance	0.7623						
13. Startle response	0.8162						

TABLE 5. The items of the Standard for Clinicians' Interview in Psychiatry (SCIP) core depression scale, item rest correlation, mean interitem correlation, Cronbach's alpha with one-sided 95% confidence interval (CI), sensitivity and specificity at the optimal cutpoint, and receiver operating characteristic (ROC) area with standard error (SE). Data based upon 700 patients interviewed at Sharpe Hospital and Chestnut Ridge Center.

CORE DEPRESSION SCALE ITEMS	ITEM REST CORRELATION	MEAN INTERITEM CORRELATION	CRONBACH'S ALPHA (ONE-SIDED 95% CI)	VALIDITY AT CUTPOINT	SENSITIVITY	SPECIFICITY	ROC AREA (SE)
1. Depressed mood	0.840	0.563	0.912 (≥ 0.903)	≥ 6	93.24%	72.32%	0.8481 (0.0151)
2. Anhedonia	0.817						
3. Hopelessness	0.825						
4. Diminished concentration	0.780						
5. Psychomotor retardation	0.693						
6. Worthlessness	0.786						
7. Guilt	0.668						
8. Suicide	0.325						

21. Hardcastle D. A follow-up study of one hundred cases made for the department of psychological medicine, Guy's hospital. *Journal of Mental Science*. 1934;80:536–549.
22. Lewis A. Melancholia: A clinical survey of depressive states. *Journal of Mental Science*. 1934;80:277–378.
23. Kendler KS. The phenomenology of major depression and the representativeness and nature of DSM criteria. *Am J Psychiatry*. 2016;173:771–780.
24. Scott K, Lewis CC. Using measurement-based care to enhance any treatment. *Cogn Behav Pract*. 2015;22:49–59.
25. Lambert MJ. Psychotherapy outcome and quality improvement: introduction to the special section on patient-focused research. *J Consult Clin Psychol*. 2001;69:147–149.
26. Hatfield DR, Ogles BM. Why some clinicians use outcome measures and others do not. *Adm Policy Ment Health*. 2007;34:283–291.

TABLE 6. The items of the Standard for Clinicians' Interview in Psychiatry (SCIP) core mania scale, item rest correlation, mean interitem correlation, Cronbach's alpha with one-sided 95% confidence interval (CI), sensitivity and specificity at the optimal cutpoint and receiver operating characteristic (ROC) area with standard error (SE). Data based upon 700 patients interviewed at Sharpe Hospital and Chestnut Ridge Center.

CORE MANIA SCALE ITEMS	ITEM REST CORRELATION	MEAN INTERITEM CORRELATION	CRONBACH'S ALPHA (ONE-SIDED 95% CI)	VALIDITY AT CUTPOINT	SENSITIVITY	SPECIFICITY	ROC AREA (SE)
1. Elated mood	0.6063	0.4855	0.9042 (≥ 0.8951)	≥ 4	95.12%	79.93%	0.9160 (0.0110)
2. Irritable mood	0.6301						
3. Mixed mood	0.3557						
4. Racing thoughts	0.7698						
5. Pressured speech	0.7450						
6. Distraction	0.7020						
7. Over activities	0.7982						
8. Grandiosity	0.5279						
9. Over spending	0.7661						
10. Decreased sleep	0.7125						

TABLE 7. The items of the Standard for Clinicians' Interview in Psychiatry (SCIP) core schizophrenia scale, item rest correlation, mean interitem correlation, Cronbach's alpha with one-sided 95% confidence interval (CI), sensitivity and specificity at the optimal cutpoint and receiver operating characteristic (ROC) area with standard error (SE). Data based upon 700 patients interviewed at Sharpe Hospital and Chestnut Ridge Center.

CORE SCHIZOPHRENIA SCALE ITEMS	ITEM REST CORRELATION	MEAN INTERITEM CORRELATION	CRONBACH'S ALPHA (ONE-SIDED 95% CI)	VALIDITY AT CUTPOINT	SENSITIVITY	SPECIFICITY	ROC AREA (SE)
1. Hallucination quality	0.6613	0.2154	0.8317 (≥ 0.8141)	≥ 2	90.12%	89.39%	0.9265 (0.0150)
2. Hallucination frequency	0.6689						
3. Hallucination duration	0.6567						
4. Voices commenting	0.5977						
5. Visual hallucination	0.5415						
6. Other hallucinations	0.1696						
7. Thought insertion	0.5702						
8. Thought withdrawal	0.3182						
9. Thought broadcast	0.4717						
10. Paranoid delusions	0.5995						
11. Conspiracy delusion	0.4778						
12. Delusion of reference	0.3779						
13. Other delusion	0.1106						
14. Bizarreness of delusion	0.3817						
15. Derailment	0.2916						
16. Tangentiality	0.2820						
17. Incoherent speech	0.1908						
18. Other disorganizations	0.2579						

TABLE 8. The items of the Standard for Clinicians' Interview in Psychiatry (SCIP) core alcohol scale, item rest correlation, mean interitem correlation, Cronbach's alpha with one-sided 95% confidence interval (CI), sensitivity and specificity at the optimal cutpoint and receiver operating characteristic (ROC) area with standard error (SE). Data based upon 700 patients interviewed at Sharpe Hospital and Chestnut Ridge Center.

CORE ALCOHOL SCALE ITEMS	ITEM REST CORRELATION	MEAN INTERITEM CORRELATION	CRONBACH'S ALPHA (ONE-SIDED 95% CI)	VALIDITY AT CUTPOINT	SENSITIVITY	SPECIFICITY	ROC AREA (SE)
1. Tolerance	0.6932	0.5828	0.9072 (≥ 0.8981)	≥ 2	79.31%	97.10%	0.9391 (0.0111)
2. Withdrawal	0.7044						
3. Failure of obligations	0.7750						
4. Social problems	0.5997						
5. Alcohol with a problem	0.8431						
6. Alcohol with hazard	0.6499						
7. Blackout	0.7776						

TABLE 9. The items of the Standard for Clinicians' Interview in Psychiatry (SCIP) core drug scale, item rest correlation, mean interitem correlation, Cronbach's alpha with one-sided 95% confidence interval (CI), sensitivity and specificity at the optimal cutpoint, and receiver operating characteristic (ROC) area with standard error (SE). Data based upon 700 patients interviewed at Sharpe Hospital and Chestnut Ridge Center.

CORE DRUG SCALE ITEMS	ITEM REST CORRELATION	MEAN INTERITEM CORRELATION	CRONBACH'S ALPHA (ONE-SIDED 95% CI)	VALIDITY AT CUTPOINT	SENSITIVITY	SPECIFICITY	ROC AREA (SE)
1. Tolerance	0.7343	0.5324	0.8723 (≥ 0.8596)	≥ 2	59.65%	91.54%	0.8515 (0.0168)
2. Withdrawal	0.7095						
3. Failure of obligations	0.7384						
4. Social problems	0.4353						
5. Drug with a problem	0.8030						
6. Drug with hazard	0.6279						

TABLE 10. The items of the Standard for Clinicians' Interview in Psychiatry (SCIP) core adult attention deficit hyperactivity disorder (ADHD) scale, item rest correlation, mean interitem correlation, Cronbach's alpha with one-sided 95% confidence interval (CI), sensitivity and specificity at the optimal cutpoint and receiver operating characteristic (ROC) area with standard error (SE). Data based upon 40 patients interviewed at Sharpe Hospital and Chestnut Ridge Center.

CORE ADULT ADHD SCALE ITEMS	ITEM REST CORRELATION	MEAN INTERITEM CORRELATION	CRONBACH'S ALPHA (ONE-SIDED 95% CI)	VALIDITY AT CUTPOINT	SENSITIVITY	SPECIFICITY	ROC AREA (SE)
1. Attention difficulty	0.3670	0.2666	0.7843 (≥ 0.6864)	≥ 5	94.74%	83.33%	0.9591 (0.0264)
2. Long attention difficulty	0.4167						
3. Attention when spoken to	0.5383						
4. Changing activities	0.4024						
5. Distraction	0.5029						
6. Fidgety	0.4156						
7. Leaving seats	0.5507						
8. Restless and moving	0.4901						
9. Over activities	0.3889						
10. Impulsivity	0.4640						

27. Roy-Byrne P, Craske MG, Sullivan G, et al. Delivery of evidence-based treatment for multiple anxiety disorders in primary care: a randomized controlled trial. *JAMA*. 2010;303:1921–1928.
28. Simon GE, Ludman EJ, Bauer MS, et al. Long-term effectiveness and cost of a systematic care program for bipolar disorder. *Arch Gen Psychiatry*. 2006;63:500–508.
29. Bauer MS, McBride L, Williford WO, et al. Collaborative care for bipolar disorder: Part II. Impact on clinical outcome, function, and costs. *Psychiatr Serv*. 2006;57:937–945.
30. Hatfield D, McCullough L, Frantz SH, Krieger K. Do we know when our clients get worse? an investigation of therapists' ability to detect negative client change. *Clin Psychol Psychother*. 2010;17:25–32.
31. Hannan C, Lambert MJ, Harmon C, et al. A lab test and algorithms for identifying clients at risk for treatment failure. *J Clin Psychol*. 2005;61:155–163.
32. Katon WJ, Lin EH, Von Korff M, et al.

TABLE 11. The items of the Standard for Clinicians' Interview in Psychiatry (SCIP) aggression scale, item rest correlation, mean interitem correlation, and Cronbach's alpha with one-sided 95% confidence interval (CI). Data based upon 700 patients interviewed at Sharpe Hospital and Chestnut Ridge Center.

AGGRESSION SCALE ITEMS	ITEM REST CORRELATION	MEAN INTERITEM CORRELATION	CRONBACH'S ALPHA (ONE-SIDED 95% CI)
1. Agitation	0.4046	0.2742	0.6939 (≥ 0.6635)
2. Violence	0.5073		
3. Violence a day	0.3810		
4. Violence a period	0.3818		
5. Odd behavior	0.5514		
6. Inappropriate affect	0.3251		

TABLE 12. The items of the Standard for Clinicians' Interview in Psychiatry (SCIP) negativity scale, item rest correlation, mean interitem correlation and Cronbach's alpha with one-sided 95% confidence interval (CI). Data based upon 700 patients interviewed at Sharpe Hospital and Chestnut Ridge Center.

NEGATIVE SYMPTOM SCALE ITEMS	ITEM REST CORRELATION	MEAN INTERITEM CORRELATION	CRONBACH'S ALPHA (ONE-SIDED 95% CI)
1. Blunted affect	0.6847	0.4877	0.8264 (≥ 0.8087)
2. Avolition	0.5682		
3. Alogia	0.6744		
4. Psychomotor slowing	0.6096		
5. Poor self-care	0.5742		

TABLE 13. The items of the Standard for Clinicians' Interview in Psychiatry (SCIP) anorexia scale, item rest correlation, mean interitem correlation, and Cronbach's alpha with one-sided 95% confidence interval (CI). Data based upon 40 patients interviewed at Sharpe Hospital and Chestnut Ridge Center.

ANOREXIA SCALE ITEMS	ITEM REST CORRELATION	MEAN INTERITEM CORRELATION	CRONBACH'S ALPHA (ONE-SIDED 95% CI)
1. Very thin	0.4009	0.2496	0.7496 (≥ 0.6398)
2. Weight affect feeling	0.3134		
3. Fear of weight gain	0.5464		
4. Losing weight by fasting	0.6139		
5. Losing weight by exercise	0.2711		
6. Losing weight by diet pills	0.3373		
7. Losing weight by vomiting	0.5962		
8. Losing weight by laxatives	0.4392		
9. Other losing weight methods	0.3417		

TABLE 14. The items of the Standard for Clinicians' Interview in Psychiatry (SCIP) anorexia scale, item rest correlation, mean interitem correlation, and Cronbach's alpha with one-sided 95% confidence interval (CI). Data based upon 40 patients interviewed at Sharpe Hospital and Chestnut Ridge Center.

BINGE EATING SCALE ITEMS	ITEM REST CORRELATION	MEAN INTERITEM CORRELATION	CRONBACH'S ALPHA (ONE-SIDED 95% CI)
1. Binge eating	0.9585	0.7434	0.9666 (≥ 0.9521)
2. Binge eating frequency	0.9366		
3. Losing control	0.7628		
4. Binge eating behavior	0.9585		
5. Eat fast	0.7743		
6. Eat until full	0.9315		
7. Eat when not hungry	0.8544		
8. Eat alone	0.6485		
9. Feel disgusted/guilty	0.7765		
10. Distressed by overeating	0.8714		

TABLE 15. The items of the Standard for Clinicians' Interview in Psychiatry (SCIP) bulimia scale, item rest correlation, mean interitem correlation and Cronbach's alpha with one-sided 95% confidence interval (CI). Data based upon 40 patients interviewed at Sharpe Hospital and Chestnut Ridge Center.

BULIMIA SCALE ITEMS	ITEM REST CORRELATION	MEAN INTERITEM CORRELATION	CRONBACH'S ALPHA (ONE-SIDED 95% CI)
1. Binge eating	0.9187	0.6088	0.9655 (≥ 0.9511)
2. Binge eating frequency	0.9437		
3. Losing control	0.7330		
4. Binge eating behavior	0.9187		
5. Eat fast	0.7264		
6. Eat until full	0.9098		
7. Eat when not hungry	0.8223		
8. Eat alone	0.6721		
9. Feel disgusted/guilty	0.7574		
10. Distressed by overeating	0.8894		
11. Compensatory behavior	0.9190		
12. Losing weight by fasting	0.5864		
13. Losing weight by exercise	0.6744		
14. Losing weight by diet pills	0.6407		
15. Losing weight by vomiting	0.5884		
16. Losing weight by laxatives	0.5864		
17. Other losing weight methods	0.5817		
18. Compensatory behavior frequency	0.9418		

Collaborative care for patients with depression and chronic illnesses. *N Engl J Med.* 2010;363:2611–2620.

33. Sapyta J, Riemer M, Bickman L. Feedback to clinicians: theory, research, and practice. *J Clin Psychol.* 2005;61:145–153.
34. Trivedi MH, Daly EJ. Measurement-based care for refractory depression: a clinical decision support model for clinical research and practice. *Drug Alcohol Depend.* 2007;88 Suppl 2:S61–71.
35. Unutzer J, Chan YF, Hafer E, et al. Quality improvement with pay-for-performance incentives in integrated behavioral health care. *American journal of public health and the nation's health.* 2012;102:e41–45.
36. Pomerantz AS, Kearney LK, Wray LO, et al. Mental health services in the medical home in the Department of Veterans Affairs: factors for successful integration. *Psychol Serv.* 2014;11:243–253.
37. Johnson-Lawrence V, Zivin K, Szymanski BR, et al. VA primary care-mental health integration: patient characteristics and receipt of mental health services, 2008–2010. *Psychiatr Serv.* 2012;63:1137–1141.
38. Sachs GS. Strategies for improving treatment of bipolar disorder: integration of measurement and management. *Acta Psychiatr Scand Suppl.* 2004;7–17.
39. Hatfield DO, B. The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice.* 2004;35:485–491.
40. WHO. *Schedules for Clinical Assessment in Neuropsychiatry Manual.* Geneva, World Health Organization; 1994.
41. WHO. *Schedules for Clinical Assessment in Neuropsychiatry Glossary.* Geneva, World Health Organization; 1994.
42. WHO. *Composite International Diagnostic Interview (CIDI).* Geneva World Health Organization; 1990.
43. Wing JK, Babor T, Brugha T, et al. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Arch Gen Psychiatry.* 1990;47:589–593.
44. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry.* 1998;59(Suppl 20):22–33;quiz 34–57.

APPENDIX I. Standard for Clinicians' Interview in Psychiatry (SCIP) depression dimension and scale

CODES

Unless otherwise specified in the question, the rating of a symptom is as follows:

0=Absent or non-significant

1=Symptom present <50% of the time or <50% of times

2=Symptom present >50% of the time or >50% of times

A positive rating of 1 or 2 implies that the patient has the symptom more than most people, or has at least some distress, or seeks professional help.

Questions apply to the present episode, typically the past month, unless otherwise specified.

Items in bold make up the core depression scale.

MB1. Depressed mood: Have you been feeling sad, depressed, or in low spirits?

- 0 Patient
- 1 Patient has depressed mood less than half the time
- 2 Patient has depressed mood more than half the time

MB2. Anhedonia: Have you been unable to experience pleasure and enjoy things that you used to enjoy, such as exercising, enjoying your hobbies, or socializing with friends?

- 0 Patient has no anhedonia
- 1 Patient has anhedonia less than half the time
- 2 Patient has anhedonia more than half the time

MB3. Crying when depressed: Have you cried when depressed?

- 0 Patient has no crying spells
- 1 Patient has crying spells due to sadness less than half the time
- 2 Patient has crying spells due to sadness more than half the time

MB4. Hopelessness: Have you felt hopeless about your future?

- 0 Patient is not hopeless
- 1 Patient feels hopeless less than half the time
- 2 Patient feels hopeless more than half the time

MB5. Diminished concentration: Have you found that your concentration has decreased and you are unable to complete a task (e.g., at work, reading an article, reading a book, or watching a movie), even though you were able to do that before?

- 0 Patient has no concentration problems
- 1 Patient has difficulty concentrating less than half the time
- 2 Patient has difficulty concentrating more than half the time

MB6. Psychomotor slowing: Have you felt as though you were talking or moving more slowly than normal for you when depressed?

- 0 Patient has normal energy and activity
- 1 Patient has psychomotor retardation less than half the time
- 2 Patient has psychomotor retardation more than half the time

MB7A: Poor appetite: Have you lost your appetite recently?

- 0 Patient has no loss of appetite
- 1 Patient had marked loss of appetite for 2 weeks or less
- 2 Patient had marked loss of appetite for more than 2 weeks

MB7B: Increased appetite: Has your appetite increased recently?

- 0 Patient had no increase of appetite
- 1 Patient had marked increase of appetite for 2 weeks or less
- 2 Patient had marked increase of appetite for more than 2 weeks

MB8: Weight loss: Did you lose weight?

- 0 Patient had no weight loss or minimal weight loss
- 1 Patient lost more than 5% of body weight in a month
- 2 Patient lost more than 15% of body weight in a year

MB9: Weight gain: Did you gain weight?

- 0 Patient had no weight gain or minimal weight gain
- 1 Patient gained more than 5% of body weight in a month
- 2 Patient gained more than 15% of body weight in a year

MB10. Sleeping problems: Have you had sleeping problems when depressed?

- 0 Patient has no sleeping problems
- 1 Patient has difficulty falling asleep (one hour or more) more than half the time when depressed
- 2 Patient has difficulty staying asleep (awakens and stays awake one hour or more) more than half the time when depressed
- 3 Patient has both difficulty falling asleep and difficulty staying asleep more than half the time when depressed

MB11. Hypersomnia: Have you been sleeping a lot more than usual when depressed?

- 0 Patient has no hypersomnia
- 1 Patient has excessive sleep (sleeps longer than 12 hours in a 24-hour period, including naps) more than half the time

MB12. Loss of libido: Has your interest in sex or your sexual activity been less than usual when depressed?

- 0 Patient has no change in sexual activities or interest in sex
- 1 Patient has much lower or no interest in sex or sexual activities

MB13. Feeling worthless: Have you felt that you are a worthless person in society or a failure?

- 0 Patient has no feeling of worthlessness
- 1 Patient feels worthless less than half the time
- 2 Patient feels worthless more than half the time

MB14. Excessive guilt: Have you felt guilty or ashamed of yourself for something you have done or thought?

- 0 Patient has no feeling of guilt
- 1 Patient feels guilty less than half the time
- 2 Patient feels guilty more than half the time

MB15A. Suicidal ideation: During the past month, have you had thoughts about harming yourself?

- 0 Patient had no suicidal ideation
- 1 Patient had suicidal ideation

MB15B. Suicidal intention: Have you had the intention to carry out the suicidal thoughts?

- 0 Patient had no suicidal intention
- 1 Patient had suicidal intention

MB15C. Suicidal plan: Have you had plans to harm yourself?

- 0 Patient had no suicidal plans
- 1 Patient had suicidal plans

MB15D. Suicidal attempt: Have you made a suicide attempt recently?

- 0 Patient made no suicide attempt during the past month
- 1 Patient made one recent suicide attempt during the past month
- 2 Patient made two or more recent suicide attempts during the past month

45. Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. *Arch Gen Psychiatry*. 1992;49:624–629.
46. Williams JB, Gibbon M, First MB, et al. The Structured Clinical Interview for DSM-III-R (SCID). II. Multisite test-retest reliability. *Arch Gen Psychiatry*. 1992;49:630–636.
47. Aboraya A. The validity results of the Standard for Clinicians' Interview in Psychiatry (SCIP). *Schizophrenia Bull*. 2015;41:S103–S104.
48. Aboraya A, El-Missiry A, Barlowe J, et al. The reliability of the standard for clinicians' interview in psychiatry (SCIP): a clinician-administered tool with categorical, dimensional and numeric output. *Schizophr Res*. 2014;156:174–183.
49. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23:56–62.
50. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13:261–276.
51. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale (BPRS): recent developments in ascertainment and scaling. *Psychopharmacol Bull*. 1988;24:97–99.
52. Goodman WK, Price LH, Rasmussen SA, et al. The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. *Arch Gen Psychiatry*. 1989;46:1006–1011.
53. Beck AT, Ward CH, Mendelson M, et al. An inventory for measuring depression. *Arch Gen Psychiatry*. 1961;4:561–571.
54. Zimmerman M, Posternak MA, McGlinchey J, et al. Validity of a self-report depression symptom scale for identifying remission in depressed outpatients. *Compr Psychiatry*. 2006;47:185–188.
55. Zimmerman M, McGlinchey JB. Why don't psychiatrists use scales to measure outcome when treating depressed patients? *J Clin Psychiatry*. 2008;69:1916–1919.
56. Nasrallah H. Long overdue: measurement-based psychiatric practice. *Current Psychiatry*. 2009;8:14–16.
57. Zimmerman M. Measuring outcome in clinical practice. *Psychiatric Times*. 2014;31:28.
58. Duffy FF, Chung H, Trivedi M, et al. Systematic use of patient-rated depression severity monitoring: is it helpful and feasible in clinical psychiatry? *Psychiatr Serv*. 2008;59:1148–1154.
59. Gilbody SM, House AO, Sheldon TA. Psychiatrists in the UK do not use outcomes measures. National survey. *Br J Psychiatry*. 2002;180:101–103.
60. Bastiaens L. Poor practice, managed care, and magic pills: have we created a mental health monster? *Psychiatric Times*. 2011;28:1–4.
61. Aboraya A. Use of structured interviews by psychiatrists in real clinical settings: results of an open-question survey. *Psychiatry (Edmont)*. 2009;6:24–28.
62. Rettew DC, Lynch AD, Achenbach TM, et al. Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *Int J Methods Psychiatr Res*. 2009;18:169–184.
63. Busner J, Kaplan SL, Greco NT, Sheehan DV. The use of research measures in adult clinical practice. *Innov Clin Neurosci*. 2011;8:19–23.
64. Aboraya A, Nasrallah H, Muvvala S, et al. The Standard for Clinicians' Interview in Psychiatry (SCIP): a clinician-administered tool with categorical, dimensional, and numeric output-conceptual development, design, and description of the SCIP. *Innov Clin Neurosci*. 2016;13:31–77.
65. Aboraya A, Nasrallah HA. Perspectives on the Positive and Negative Syndrome Scale (PANSS): Use, misuse, drawbacks, and a new alternative for schizophrenia research. *Ann Clin Psychiatry*. 2016;28:125–131.
66. Arbuckle MR, Weinberg M, Harding KJ, et al. The feasibility of standardized patient assessments as a best practice in an academic training program. *Psychiatr Serv*. 2013;64:209–211.
67. Roy-Byrne P, Dagadakis C, Ries R, et al. A psychiatrist-rated battery of measures for assessing the clinical status of psychiatric inpatients. *Psychiatr Serv*. 1995;46:347–352.
68. Zimmerman M, McGlinchey JB. Depressed patients' acceptability of the use of self-administered scales to measure outcome in clinical practice. *Ann Clin Psychiatry*. 2008;20:125–129.
69. Baer L. *Handbook of Clinical Rating Scales and Assessment in Psychiatry and Mental Health*. Boston: Humana Press; 2015.
70. Andreasen NC, Flaum M, Arndt S. The Comprehensive Assessment of Symptoms and History (CASH). An instrument for assessing diagnosis and psychopathology. *Arch Gen Psychiatry*. 1992;49:615–623.
71. Blumenthal D. Stimulating the adoption of health information technology. *N Engl J Med*. 2009;360:1477–1479.
72. Long N. Closing the gap between research and practice—the importance of practitioner training. *Clin Child Psychol Psychiatry*. 2008;13:187–190.
73. Baker M. Measurement-based care and psychiatric education: are we doing enough? *Psychiatric News*. 2015;50.
74. The Psychiatry Milestone Project Initiative by the Accreditation Council for Graduate Medical Education and American Board of Psychiatry and Neurology. 2015. <https://www.acgme.org/Portals/0/PDFs/Milestones/PsychiatryMilestones.pdf>. Accessed 1 Dec 2018.
75. Spitzer RL, Fleiss JL, Burdick EI, Hardesty AS. The mental status schedule: rationale, reliability and validity. *Compr Psychiatry*. 1964;5:384–395.
76. Spitzer RL. Psychiatric diagnosis: are clinicians still necessary? *Compr Psychiatry*. 1983;24:399–411.
77. Valenstein M, Adler DA, Berlant J, et al. Implementing standardized assessments in clinical care: now's the time. *Psychiatr Serv*. 2009;60:1372–1375. **ICNS**